Text-mining Techniques and Tools for Systematic Literature Reviews: A Systematic Literature Review

Luyi Feng
Faculty of Computer Science and Information Technology
University of Malaya
Kuala Lumpur, Malaysia
Email: janqualine@163.com

Yin Kia Chiam
Faculty of Computer Science and Information Technology
University of Malaya
Kuala Lumpur, Malaysia
Email: yinkia@um.edu.my

Sin Kuang Lo
Faculty of Computer Science and Information Technology
University of Malaya
Kuala Lumpur, Malaysia
Email: dnllo36@gmail.com

***Abstract*—Despite the importance of conducting systematic literature reviews (SLRs) for identifying the research gaps in software engineering (SE) research, SLRs are a complex, multi-stage, and time-consuming process if performed manually. Conducting an SLR in line with the guidelines and practice in the SE domain requires considerable effort and expertise. The objective of this SLR is to identify and classify text-mining techniques and tools that can help facilitate SLR activities. This study also investigates the adoption of text-mining (TM) techniques to support SLR in the SE domain. We performed a mixed search strategy to identify relevant studies published from January 1, 2004, to December 31, 2016. We shortlisted 32 papers into the final set of relevant studies published in the SE, medicine and social science disciplines. The majority of the text-mining techniques attempted to support the study selection stage. Only 12 out of the 14 studies in the SE domain applied text-mining techniques, focusing primarily on facilitating the search and study selection stages. By learning from the experience of applying TM techniques in clinical medicine and social science fields, we believe that SE researchers can adopt appropriate SLR automation strategies for use in the SE field.**

***Index Terms*—Text mining techniques; Systematic Literature Review; Tool support**

I. Introduction

The evidence-based paradigm and practice was first developed in the clinical medicine area. It was extended to other domains and is prevalent across many disciplines, including SE. One of the core tools of evidence-based software engineering (EBSE) is the systematic literature review (SLR). SLR is a well-established research method used to integrate the best available empirical data from systematic research [1], [2]. In contrast to the conventional ad hoc literature review process, SLR provides reliable approaches and established sequences to aggregate, evaluate and interpret the best available individual studies to address particular research questions [2]. It also allows reviewers to determine the real effects and phenomena in areas where small, individual studies are not easily controlled or replicated [3].

Despite its usefulness and popularity, a SLR is a challenging task to perform manually. Due to its complex nature, the SLR process is time consuming, labor intensive and prone to errors [3], [4], [5], [6], [7], [8]. Unlike conventional literature review procedures, SLRs insist on the use of comprehensive and rigorous search strategies to retrieve as many relevant studies as possible. Consequently, they typically yield thousands or millions of results, and a majority of such results are irrelevant studies. Because any failure to include eligible primary studies will threaten the validity of the review, the task of study selection for SLR requires great patience and domain expertise. As a result, screening such a large volume of citations is an arduous process. This also leads to difficulties in performing the remaining steps of the SLR process.

The increase number of publications in SE has made the task of identifying and selecting relevant articles in an unbiased way for SLRs becomes more time consuming and tough [3], [5], [6]. Additionally, after selecting the primary studies, SE researchers need to screen and analyse the selected papers to get valuable results through data synthesis. To address these issues, several SE studies have explored the potential of applying TM techniques such as Visual Text Mining (VTM) [9], [10], [11], classification methods[12] and Vector Model[13] and develop tools based on TM techniques to facilitate SLR processes since 2007. TM is the process of deriving interesting information, patterns, non-trivial knowledge and trends from textual documents [14]. Despite the fact that SLRs have been introduced to SE field for more than a decade, the prospect of having good supporting tools for SLRs, built based on TM techniques is promising but the effectiveness of using TM techniques have not fully explored and tested.

This study presents an SLR in an attempt to understand the application of different TM techniques in facilitating the SLR process. We are interested in identifying the main challenges in the SLRs that can be addressed by applying TM techniques. SLR activities that can be supported by TM techniques and the extent to which these activities can be automated. Additionally, we would like to investigate the issues faced by the researchers when applying TM techniques to support SLR in the SE domain. We believe that this review will benefit researchers who want to explore the application of TM techniques to improve the SLR process in the SE domain.

The remainder of this paper is organized as follows: Section II provides an overview of TM techniques and previous review related to this study. Section III describes the review method, and the SLR results are presented in Section IV. Section V uses the data extracted from identified evidence to

978-1-5386-3681-7/17 $31.00 © 2017 IEEE
DOI 10.1109/APSEC.2017.10

answer the review questions. Section VI concludes the study.

II. Background

This section presents the background of TM techniques and discusses previous studies that have conducted a similar systematic literature review.

A. Text mining techniques

Text mining (TM), also referred to as intelligent text analytics, text data mining and knowledge discovery in text (KDT), is the process of deriving useful information from text. The process often involves various techniques, including information retrieval, data mining, machine learning, natural language processing (NLP), and knowledge management [15].

Various studies [16], [17], [18] have discussed applications of TM techniques. Based on the previous studies, we identified the basic TM techniques and classified them into six main categories. Table I presents descriptions of the six categories of TM techniques: information extraction (IE), information retrieval (IR), information visualization (IVi), document classification, document clustering and document summarization. In this SLR, we explored how the different types of TM techniques have been applied to facilitate the SLR process.

B. Previous Review

Several reviews for the SLR process have been reported in the last few years. Felizardo et al. [10] conducted a mapping study on the use of visual data-mining (VDM) techniques to support the SLR process. The authors studied 20 primary studies, and the results show a scarcity of research on the use of VDM to help with performing SLRs in the SE field. According the review results, only one of 20 primary studies adopted VDM in the SE domain. This demonstrates that there is a research gap between the SE and ME domains; however, it does not explain the reason for this research disparity. The mapping study also indicates that existing studies have applied only VDM techniques in the process of search and study selection. Future research is needed to develop new VDM techniques and tools to support other SLR phases or activities, particularly in the planning and reporting phases. In this SLR, we will extend the review to investing how TM techniques are used to support SLR processes.

III. Review method

This SLR was conducted based on the guidelines proposed by Kitchenham [1], [2]. According to the guidelines, the SLR process starts with the formulation of review questions and continues with the development and validation of a review protocol, followed by the search for and screening of primary studies based on the criteria defined in the review protocol. Next, we accessed the full text of identified studies and adopted a collaborative review method to assess their quality. By analyzing the data and undertaking the data synthesis, we extracted final results from the review. The following subsections present the details of our review method.

A. Research questions

The objective of this paper is to investigate TM techniques that can be used to support SE researchers in the SLR process. To this end, we defined a set of research questions (RQs) to be addressed in this SLR.

The research questions for this review are as follows:

- RQ1: What TM techniques have been applied to facilitate the SLR process?
- RQ2: How do TM techniques support SLR activities?
- RQ3: How will the adoption of TM techniques facilitate SLRs in the SE domain?

B. Review Protocol

We developed a review protocol for this SLR according to the SLR guidelines proposed by Kitchenham [1], [2]. This protocol includes the search strategy, inclusion and exclusion criteria, quality assessment, data extraction, and data synthesis.

1) Search Strategy: To address the research questions, we performed automated search strategy. We adapted the syntax and search algorithms of different digital libraries to define the query string for database searching. The query strings and search scope are as follows:

- Query String - ("systematic literature review" OR "systematic review" OR "evidence-based") AND (clustering OR classification OR "information extraction" OR "information retrieval" OR "information visualization" OR "summarization" OR "text mining")
- Timespan: January 1, 2004 to December 31, 2014
- Language: English
- Reference Type: Journal, Conference, Workshop, Symposium, Book Chapter

We used Google Scholar as a retrieval facility to conduct snowballing in our search process to identify new papers [19]. The papers that fulfilled the inclusion and exclusion criteria were selected for this review. The initial timespan of this SLR is 2004 to 2014, to extend this SLR to cover 2015 and 2016 papers, we reached a consensus to perform a manual search of software engineering conference proceedings and journals.

2) Inclusion and exclusion criteria: In this SLR, we defined a set of inclusion and exclusion criteria based on the research questions to shortlist the relevant papers.

Studies that fulfilled one of the following inclusion criteria were shortlisted:

- The study either described the use of a tool or applied at least one TM technique to support SLR process or activities.
- If the paper was a review paper, we identified the studies reported in the review and separately treated each paper.
- The most recent version of the study was included if several papers reported the same study.

Studies that fulfilled the following criteria were excluded:

- The study reported the SLR but did not use a tool or apply TM techniques to support the SLR.

TABLE I
CATEGORIES OF TM TECHNIQUES

TM Category	Description
Information Extraction (IE)	Finding a specific piece of information from a text document using a pattern-matching method to find key phrases and relationships with the text.
Information Retrieval(IR)	Investigation of appropriate mechanisms for searching interesting information from a collection of resources.
Information Visualization (IVi)	Visualize information using interactive visual representations to amplify human recognition.
Classification (Categorization)	Finding interesting patterns/features that make them part of a defined grouping and assigning documents to known categories.
Clustering	Finding interesting patterns associated with extracted data and grouping similar documents based on their content.
Summarization	Reducing the length and detail of the source text into a shorter version while preserving the implied meaning of its information.

- The study described the theoretical aspects of performing SLR activities (e.g., guidelines) or proposed manual techniques to conduct SLR but did not apply TM techniques to automate the SLR activities.
- The study was a duplicate.
- The study was not written in English.

The study selection process was performed in three phases. First, the first author conducted the search and found all potential primary studies. Next, the first two authors by reading the title and abstract of all papers obtained from the search results. The studies that fulfilled the inclusion and exclusion criteria were selected. Next, all of the authors read the selected papers in full to shortlist the papers for final selection.

3) Quality assessment: During this stage, all of the authors assessed the quality of the research presented by the studies that had fulfilled the inclusion and exclusion criteria. To analyze the rigorousness, credibility and relevance of the studies, we independently evaluated each study according to 10 criteria. For each paper, we read the full text and applied the assessment criteria to assess its quality. As shown in Table II, the criteria were a set of quality assessment questions developed based on the checklists by Dybå, Tore and Dingsøyr [20], and Nguyen-Duc et al. [21]. Each question had four possible scores: completely addressed (3), adequately addressed (2), little mentioned (1) and not mentioned at all (0). All primary studies were sorted according to their score.

4) Data extraction: During this stage, we collected all of the information needed to perform an in-depth analysis to address all of the research questions. We created a predefined extraction form to record the following data extracted from all of the selected primary studies.

- Reference ID
- Reference bibliographic information
- Reference type
- Publication name
- Institution country
- Research discipline
- Type of TM techniques applied in the SLR process
- Underlying algorithms/models/theories
- Specific SLR stage that has been supported
- Tool(s) that has been used in the studies to facilitate the SLR process

5) Data Analysis: After extracting the information from each primary study, we conducted an in-depth data analysis

TABLE II
QUALITY ASSESSMENT CRITERIA

No.	Criterion
	Problem Statement
Q1	Is the research objective sufficiently explained and well-motivated?
	Research Design
Q2	Is it clear which TM technique(s) can be used to support the SLR process?
Q3	Is it clear which SLR activities can be supported using the TM techniques or automation methodologies?
	Data Collection
Q4	Are the data collection and measures adequately described?
Q5	Are the measures and constructs used in the study the most relevant for answering the research question/issue?
	Data Analysis
Q6	Is the data analysis used in the study adequately described?
Q7a	Qualitative study: Is the interpretation of evidence clearly described?
Q7b	Quantitative study: Has the significance of the data been assessed?
Q8	Is it clear how the TM technique(s) or supporting tool(s) have been used?
	Conclusion
Q9	Are the findings of the study clearly stated and supported by the results?
Q10	Does the paper discuss the limitations or validity?

to answer each research question. For RQ1, we identified and classified the TM techniques used in the SLR for each primary study. We also analyzed the algorithms, models or theories of the underlying TM techniques in each category. For RQ2, we matched the proposed method to the relevant SLR phases or activities facilitated by applying TM techniques. The level of SLR automation in applying these TM techniques was analyzed. For RQ3, we analyzed the number of adopted TM techniques with regard to different disciplines and the distribution of the institution countries. For RQ4, the tools described in the selected studies were identified and classified. We also specified and analyzed the potential tools that can be used to support the SLR process in the SE domain.

IV. RESULTS

This section summarizes the results collected based on the method described in the previous section from this SLR.

A. Search results and selection of primary studies

The search for and selection of primary studies were conducted based on the aforementioned search strategy. Fig. 1 shows the process involved in the identification and selection of the relevant studies. First, we searched the six electronic databases for relevant studies using the query strings defined

TABLE III
SEARCH RESULTS

Digital Library	Returned Papers	Final Set
IEEE Xplore	223	7
Web of Science	184	3
SpringerLink	234	2
ScienceDirect	246	3
ACM Digital	233	4
Google Scholar	2,920	8
Manual search	16	5
Total:	4056	32

in Section III-B1 plus digital snowballing. A total of 4,040 results were retrieved from the automated searches. Then, we exported the search results into EndNote for reference management and used its "Find Duplicates" function to remove 940 identical papers. We then conducted collaborative citation screenings using EndNote. Based on the inclusion and exclusion criteria, only 51 papers were found after eliminating 3,049 irrelevant papers. Next, we obtained the full text of these 51 studies and perform snowballing manually to the reference list of the shortlisted primary studies. Additional 19 relevant papers were identified through snowballing. Finally, we conducted a peer review to analyze the full text of the selected studies, and only 27 studies out of 70 relevant papers were included in the final set of primary studies for this SLR. In the manual search, we have identified 16 papers (published in 2015 and 2016) and selected 5 relevant papers for this review. Table III shows the 32 papers that were selected after eliminating irrelevant or duplicate papers.

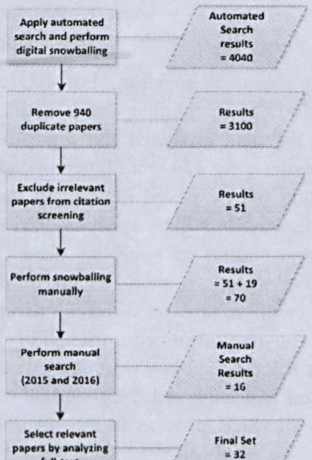

Fig. 1. Identification and selection of primary studies

Table IV shows the complete list of the 32 primary studies. These studies are journal articles, conference papers, book sections and workshop papers published in SE, ME and SS disciplines. There are 10 papers from SE, 16 papers from ME and 1 paper from SS. We used the data extraction form designed in Section III-B4 to extract data from the selected studies. In terms of country, the USA has contributed the majority of the studies (36%) and majority of them focus on building machine learning classifiers for automatic study selection. Brazil represented the second largest portion of studies (21%) and Brazilian researchers mainly focus on the development of information visualization and clustering techniques for the SLR process.

B. *Quality assessment results*

After identifying the final set of primary studies for this SLR, we assessed the quality of each selected study based on the predefined quality assessment criteria shown in Section III-B3 (see Table II). Table V shows the scores given to each study for each criterion. We only select papers with the average quality score above 1.5 to ensure the reliability of the studies. The results for the quality score in Table V show that each selected study has an average score of 2.0 or higher, which is considered fair or good quality.

V. DISCUSSION

In this section, we will answer the research questions stated in Section III-A.

A. *RQ1: What TM techniques have been applied to facilitate the SLR process?*

Overall, we identified 32 relevant studies that proposed an automated or semi-automated solution to support the SLR process. 30 papers explicitly described the use of TM techniques to facilitate SLR activities. Table VI shows the type of TM techniques and their respective studies. Various algorithms, models, and theories are used in the studies to support the SLR activities. Classification techniques are the most commonly used TM techniques reported in the studies. In addition, there are different applications of TM techniques within the existing SLR process (see Figure 2). We identified four main applications of TM techniques: 1) visual text mining (VTM), 2) federated search strategy, 3) automated document/text classification and 4) document summarization.

VTM is a type of visual data mining (VDM) technique applied to texts. It attempts to place large textual sources into a visual hierarchy or map that provides browsing capability [52], [46]. In general, VTM techniques such as document clustering and information visualization, are valuable for data cleansing. VTM allows reviewers to quickly identify potential relevant and irrelevant papers according to their similarities [9], [10], [11], [49], [50]. However, the usability of the VTM approach is questionable in the study selection stage because SLR study selection requires the selection of primary studies against inclusion or exclusion criteria. Although VTM provides a potentially useful method for mining and clustering reference information from primary studies in an SLR, VTM does not provide the holistic management of selection activities. Moreover, the selection stage consists of a two-stage selection: citation screening and full-text screening. VTM methods can support only the former stage, and the latter stage requires

TABLE IV
SELECTED PRIMARY STUDIES

ID	Author(s)	Year	Article Type	Institution Country	Discipline	Tool(s)
S01	Malheiros et al. [9]	2007	Symposium	Brazil	SE	PEx
S02	Felizardo et al. [10]	2010	Conference	Brazil	SE	PEx
S03	Ramampiaro et al. [22]	2010	Conference	Norway, Brazil	SE	MetaSearcher model
S04	Fernández-Sáez et al. [23]	2010	Conference	Spain	SE	SLR-Tool*
S05	Fabbri et al. [24]	2013	Book Chapter (Conference)	Brazil	SE	StArt*
S06	Bowes et al. [25]	2012	Workshop	Ireland, UK	SE	SLuRp*
S07	Ghafari et al. [26]	2012	Journal	Iran	SE	Federated Search tool*; Nvivo; Zotero
S08	Felizardo et al. [11]	2014	Conference	Brazil, New Zealand	SE	ReVis*
S09	Torres et al. [27]	2013	Workshop	Brazil, Norway	SE	A specific tool (using Weka library)*
S10	Aphinyanaphongs et al. [28]	2005	Journal	USA	ME	A framework for automatic generation of a gold standard
S11	Cohen et al. [29]	2006	Journal	USA	ME	EndNote; NCBI Batch; Citation Matcher
S12	Kouznetsov et al.[30]	2009	Book Chapter (Conference)	Canada	ME	N/A
S13	Martinez et al. [31]	2008	Symposium	Australia	ME	Weka
S14	Ananiadou et al. [32]	2009	Journal	UK	SS	EPPI-Reviewer
S15	Yang et al. [33]	2008	Symposium	USA	ME	SYRIAC system*
S16	Cohen et al. [34]	2009	Journal	USA	ME	N/A
S17	Matwin et al. [35]	2010	Journal	Canada	ME	JabRef
S18	Frunza et al. [36]	2010	Conference	Canada	ME	Weka
S19	Wallace et al. [37]	2010	Journal	USA	ME	Abstrackr
S20	Tomassetti [38]	2011	Conference	Italy	SE	DBPedia
S21	Cohen et al. [39]	2012	Journal	USA	ME	SYRIAC system*
S22	Shash and Mollá [40]	2013	Book Chapter (Conference)	Australia	ME	MetaMap
S23	Kim and Choi [41]	2014	Journal	Korea	ME	N/A
S24	Bekhuis et al. [42]	2014	Journal	USA	ME	EDDA*, extension with two open source plugins (from the MALLET Toolkit)
S25	Adeva et al. [43]	2014	Journal	Spain	ME	Pimiento
S26	Miwa et al. [44]	2014	Journal	UK	ME	Use LIBLINEAR library to create the classifiers
S27	Smalheiser et al. [45]	2015	Journal	USA	ME	Metta
S28	Mergel et al. [46]	2015	Conference	Brazil	SE	SLR.qub*
S29	Abilio et al. [47]	2015	Journal	Brazil	SE	Weka
S30	Piroi et al. [48]	2015	Conference	Austria	SE	DASyR*
S31	Octaviano et al. [49]	2016	Conference	Brazil	SE	StArt*, Weka
S32	Fabbri et al. [50]	2016	Conference	Brazil	SE	StArt*, Weka

NOTE: *=Specific SLR Tool developed by researchers; SE = Software Engineering, ME = Medicine, SS = Social Science; N/A = Not Available or not mentioned in the article

semantic information-extraction techniques or a document summarization system to support the full-text screening stage.

The federated search strategy attempts to provide reviewers with a unified query interface to retrieve documents from different databases by entering a single search query. The application of various IR techniques is also recommended to improve search efficiency. For example, the use of automatic query expansion techniques is an effective method for helping reviewers refine query strings. Ghafari et al. [26] first introduced the federated search approach in the SLR process. Most recently, Torres et al. [27] combined the federated search concept with ranking algorithms. The prioritization of studies enables full-text studies to be retrieved earlier in the study selection process. Smalheiser et al. [45] designed and developed a clinical search tool, called Metta, to support a clinician-conducted federated search for the SR process.

Automatic Text Classification (ATC) is often used in the early stages of search and study selection to, automatically distinguish relevant studies from irrelevant ones. There are two possible strategies for using text mining to identify relevant studies. One uses automatic term-recognition approaches [53], and the other uses classification techniques that require classifier construction (training and testing). Classification algorithms can be divided into two categories: rule-based and active (machine) learning methods. Rule-based text classification is also known as knowledge engineering approaches to achieving ATC. This technique is focused on the manual development of classification rules. Experts use their knowledge to manually define a set of rules on how to classify a document into predefined categories. However, rule-based text classification (e.g., decision tree algorithms) can be excessively tedious [30], whereas machine learning (ML) text classification concerns building ML classifiers to achieve

TABLE V
QUALITY SCORES FOR ALL SELECTED PRIMARY STUDIES

ID	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Avg
S01	3	2	3	3	3	3	3	2	3	3	2.8
S02	3	2	3	3	2	3	3	2	3	3	2.7
S03	3	3	2	2	1	2	2	1	2	2	2.0
S04	3	1	3	2	3	2	2	3	2	2	2.3
S05	3	3	3	2	2	3	2	2	2	2	2.4
S06	3	1	3	2	1	2	2	2	2	2	2.0
S07	3	2	3	2	2	2	2	1	3	3	2.3
S08	3	2	3	3	3	3	3	2	3	2	2.7
S09	3	3	3	3	3	3	3	3	3	3	3.0
S10	3	3	3	2	3	3	2	3	3	3	2.8
S11	3	3	3	3	2	3	2	2	3	3	2.7
S12	2	3	3	3	3	2	3	2	2	2	2.5
S13	3	3	3	3	2	3	3	2	3	3	2.8
S14	2	3	3	3	3	3	2	2	3	3	2.7
S15	2	3	3	3	3	3	3	3	3	3	2.9
S16	3	3	3	2	3	3	3	2	3	3	2.8
S17	3	3	3	3	3	3	3	3	3	3	3.0
S18	3	3	3	3	3	3	3	1	3	3	2.8
S19	3	3	3	3	3	3	3	3	3	3	3.0
S20	3	3	3	2	3	3	3	2	3	3	2.8
S21	2	3	3	3	3	3	3	3	3	3	2.9
S22	3	3	3	3	3	3	3	2	3	3	2.9
S23	3	3	3	3	3	3	3	3	3	3	3.0
S24	3	3	3	3	3	3	3	1	3	3	2.8
S25	3	3	3	3	3	3	3	2	3	3	2.9
S26	3	3	3	3	3	3	3	3	3	3	3.0
S27	3	3	3	2	2	3	3	2	3	2	2.6
S28	3	3	3	3	3	3	3	3	3	3	3.0
S29	2	3	3	3	3	3	3	3	3	3	2.9
S30	3	3	3	3	3	3	3	3	3	1	2.8
S31	3	3	3	3	3	3	3	3	3	3	3.0
S32	3	3	3	3	2	2	2	3	2	1	2.4

ATC. However, the performance of a classification algorithm is greatly affected by the quality of the data source.

Document summarization is the process of automatically creating a summary that retains the most important points of the original document. As the problem of information overload has grown, and as the quantity of data has increased, so has interest in automatic summarization. Generally, there are three approaches to automatic summarization: clustering methods, sentence recognition and abstraction strategy [32], [22], [40], [27]. Sentence recognition works by selecting a subset of existing words, phrases, or sentences in the original text to form the summary and using clustering techniques to identify their similarities. Research on abstractive methods is an increasingly important and active research area; however, the research to date has primarily focused on sentence-recognition methods due to complexity constraints.

B. RQ2: How do TM techniques support SLR activities?

Before answering this research question, we first turn to the following question: Which SLR activities should be automated by applying text-mining techniques? According to the experiences of performing the SLR process reported by da Silva [54], Hassler [55], and Brereton [3], we discuss the challenges faced by SE researchers and practitioners and identify the need to automate the following SLR activities.

- **Pre-review mapping:** Kitchenham et al. [2] and Brereton et al. [3] recommended conducting a pre-review mapping study during the planning phase before the development of a review protocol. However, this activity has always been ignored or skipped by SLR researchers. It is of great importance because the quality and efficiency of performing SLR greatly depend on the quality of the protocol. Applying TM to scoping activities allows the reviewer to quickly identify research topics and provide an initial overview of the existing body of knowledge. This information is useful for scoping research questions, and the identified keywords and important terms can be further used to help devise the search strings.

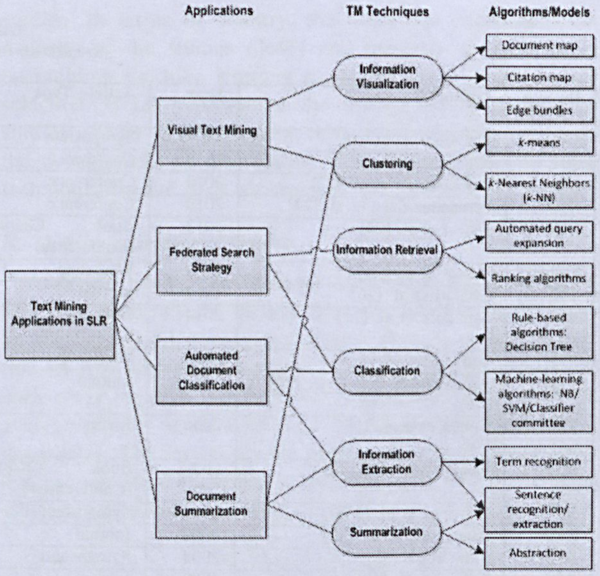

Fig. 2. Main text-mining applications in SLR

- **Search process:** A defined search strategy must be used to locate as many relevant studies as possible from different data sources. Any incomplete or inappropriate query may yield far more irrelevant studies, consequently increasing the workload of selection or even missing critically relevant studies. Worse still, because different databases have different structures, different search engines have different query syntax, and inadequate search facilities may support these electronic databases. The search process is quite time consuming and labor intensive. Current studies offer two solutions. First, Tomassetti et al. [38], Ghafari et al. [26], and Smalheiser et al. [45] agreed on the use of federated search approaches to provide a unified interface for all defined SE data sources. Reference management should be provided to manage the list of references for the searched papers. Google Scholar and Scopus provide good examples of outstanding meta-searchers, and researchers encourage more meta-searchers with good retrieval facilities to be developed in this field. Second, the search process can be improved using query-expansion techniques [28].
- **Study Selection:** An extensive and rigorous search strat-

TABLE VI

TM TECHNIQUE(S) AND ALGORITHMS, MODELS, AND THEORIES USED IN THE PROPOSED WORK

ID	TM Technique(s)	Algorithms/models/theories
S01	IVi, Clustering	VTM techniques, including clustering technique: Fastmap and ProjClus projection; Vector space model (VSM), Normalized Compression Distance (NCD) similarity calculation methods
S02	IVi, Clustering	VTM techniques, including Fastmap and ProjClus projection; VSM and NCD similarity calculation methods.
S03	IR, Clustering	Clustering algorithm: e.g., k-means or k-NN
S04	IR, clustering	Did not specify which TM techniques were used
S05	IVi and IE	Vector Processing Model [51]
S06	N/A	N/A
S07	IR	Federated search model
S08	IVi, Clustering, Summarization	VTM techniques, including KNN Edges connection techniques, HiPP clustering strategy
S09	Classification, IE	Textum approach, including both rule-based algorithms (cue methods) and ML algorithms
S10	Classification	ML classification algorithm: committee of classifiers, including CNB, MNB, alternating DT; AdaBoost (logistic regression and j48 algorithms
S11	Classification	ML classification algorithm: voting perceptron classifier; rule-based classifier (Slipper)
S12	Classification	Ranking algorithm and classification rule for the committee of classifiers included
S13	IR, Classification	Ranked queries, Regression algorithm: a version of SVM for regression, called support vector regression (SVR) for text
S14	Clustering	Clustering algorithm: lingo; information extraction algorithms: C-value, ML classification algorithm: SVM classifier
S15	Classification	ML classification algorithm: NB and SVM classifiers
S16	Classification	ML classification algorithm: SVM classifier
S17	Classification	ML classification algorithms: FCNB and FCNB/WE classifiers
S18	Classification	ML classification algorithm: CNB classifier
S19	Classification	ML classification algorithms: FCNB and FCNB/WE classifiers
S20	IR	ML classification algorithms: NB classifier
S21	classification and IE	ML classification algorithm: SVM classifier
S22	Clustering, summarization and IE	Clustering algorithm: K-means
S23	Classification	ML classification algorithm: SVM classifier
S24	Classification	ML classification algorithm: complement naive Bayes (CNB) classifier
S25	Classification, IE	ML classification algorithms: NB, SVM and Rocchio classifiers
S26	Clustering, Classification	ML classification algorithm: NB and KNN classifiers
S27	IR, IE	Federated search model
S28	IVi, IE	VTM techniques
S29	IR, Clustering	Jaccard Similarity Coefficient, Vector Model
S30	IR, Classification	Positional random indexing, ML classification algorithm: SVM and Decision Tree classifiers
S31	Classification	ML classification algorithm: Decision Tree classifiers
S32	Classification	ML classification algorithm: Decision Tree classifiers

Note: N/A = Not Available

egy typically yields thousands of documents. Therefore, the inclusion of relevant studies from a large volume of studies can be a difficult task. Existing research focuses on using different TM techniques to address this problem. Within the SE domain, researchers have proposed VTM techniques that combine document clustering and information visualization techniques [9], [10], [11]. A semi-automation study selection strategy, SCAS (Score Citation Automation Selection) was proposed by the researchers based on 50% score ranking technique and J48 decision tree technique [49], [50]. These visualized clusters allow reviewers to quickly identify associations between documents. However, these techniques require certain knowledge about TM and a thorough understanding of information visualization representations. In the ME domain, researchers used of automated document classification [28], [29], [22], [30], [55], particularly applying ML algorithms to construct classifiers. However, the performance of classification greatly relies on the quality of the primary studies.

- **Data Extraction and Synthesis**: This involves collating and summarizing the results of the included primary studies with IE techniques. It can be improved using a multi-document summarization system[40] to produce a coherent arrangement of the extracted sentences from multiple documents.
- **Updating SRs (USRs)**: This involves updating the outdated SRs when new evidence was found after the completion of previous reviews [11], [34], [46]. Updating SRs manually is a time-consuming process. Selection of new evidence or articles can be supported using VTM or ML techniques to train the predictions on the appropriateness of a new article to be included in the updated SR.
- **SLR process management**: SLR is a multi-stage collaborative review process. At least two reviewers are typically involved in performing an SLR. Therefore, tools to support collaborative review are urgently needed, especially in decision-making and reference management. Brazillian researchers[24], [49], [50] developed a tool called StArt, implemented with TM techniques to manage the SLR process.

From the collected studies, the majority of researchers focus on the search and selection stages during the conducting review phase. Through this SLR, we also identify a research gap in supporting the SLR planning and reporting review phases. Further work is needed to design new TM approaches to automate or semi-automate the planning and reporting phases. Figure 3 shows the mapping between the TM applications and SLR activities that can be automated.

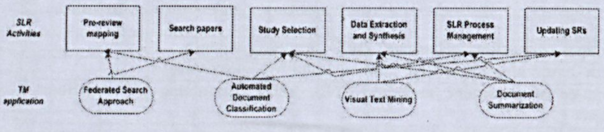

Fig. 3. Text-mining techniques and automatable SLR activities

For RQ2, we use Table VII to answer how the identified TM techniques support specific SLR activities. First, we divided overall the SLR process into 3 phases and 11 activities. We then defined sub-activities and specified the level of automation (i.e., low, medium, or high) needed to support each sub-activity. A high level indicates that the specific SLR activities in particular need TM to support automation. Further SLR automation research should focus on these highly valued SLR activities. We also analyzed the automation strategy and TM techniques that can be applied for each SLR activity.

C. RQ3: How will the adoption of TM techniques facilitate SLRs in the SE domain?

In the SE discipline, TM techniques are increasingly used to facilitate the SLR process. Brazilian researchers [9], [10], [11], [49], [50] were the first to suggest the use of the VTM strategy to support selection activity. They also introduced an existing visualization tools, called PEx, to support the study selection. Later, Felizardo et al. [11] continued research on SLR/SM-VTM to support the typical SR update process by suggesting a re-checked of primary studies and update SR with new evidence. They proposed a new VTM tool, called Reviz, to support study selection and demonstrate there is a positive relationship between the use of VTM techniques and the time required to conduct the selection review activity. Out of the 15 papers in the SE domain, 14 studies applied TM techniques to facilitate the searching, study selection and updating SRs in the SLRs. Most of the SE studies focus on applying VTM techniques to show the citation relationships. Felizardo et al. [11] proposed an USR-VTM approach that applied content map and edge bundles to provide two strategies for inclusion and exclusion of new primary studies.

SE researchers should learn from other disciplines (ME and SS), to reduce the workload associated with conducting SLRs. In other disciplines, researchers prefer to apply ML algorithms to simplify screening activities. Ananiadou et al. [32] applied an ML algorithm (i.e., support vector machines (SVMs)) to improve the screening strategy of SLRs. This work focused on improving performance over the clinical query filters first proposed by Wong et al. [56]. Cohen et al. [29] also applied ML to SLR, focusing on reducing the workload of reviewers during broad screening by eliminating as many non-relevant documents as possible while including no less than 95% of the relevant documents. They introduced a new measure: work saved over sampling (WSS). They defined work saved as "the percentage of papers that meet the original search criteria that the reviewers do not have to read (because they have been screened out by the classifier)". They then specified that the work saved must be greater than the work saved by random sampling; thus, WSS is proposed as an evaluation measure.

Different TM techniques have been applied to improve SLR processes in the ME domain[29], [33], [34], [39], [45]. The classifier performance scores that they obtained with the voting perception (VP) algorithm, the algorithm that they used for the classification task, are for the SE field. Within the ME field, the highest level of evidence is based on randomized controlled trials (RCTs) which often have unique terms and a clear classification taxonomy [57]. Moreover, there are well-built infrastructures (e.g., PubMed, and Embase) and rich resources (MeSH vocabulary) that provide a foundation for the implementation of various ML approaches to facilitate SLRs within clinical medicine research. SE researchers should consider adopting ML algorithms to facilitate SLRs in SE. However, in SE fields, context-dependent evidence is always the best evidence [58]. To overcome the problem of applying ML in the SE domain, Wohlin [58] suggested creating evidence profile for SE research and practice to support evidence synthesis and evaluation in SLRs.

VI. Conclusion

This SLR study discussed 32 studies that describe the use of TM techniques, to support the SLR process. This SLR helps us understand the state of the art in SLR automation techniques. We also identified gaps and future research directions. Performing SLR manually is labor intensive, time consuming and prone to errors. Another key challenge in SLR studies is the selection and validation of the available evidence.

The use of TM techniques and tools to support the SLR process is an active area of research. However, although many studies provide TM approaches to support the SLR process, none have applied TM techniques in the planning phase. Moreover, although the pre-review mapping study is of great importance in the SLR process, most SLR researchers have neglected this topic. In the ME and SS fields, ML techniques, such as SVMs have been widely applied to support the SLR process. Although there are some issues whereby SE research differs from other disciplines, we can learn from the experience obtained in performing SLRs in the ME and SS fields and their well-designed SLR automation strategies can be adapted for use in the SE field. The goal of the supporting tools is to encourage the widespread adoption of the SLR process. The existing tools can be improved to address the challenges of the SLR process. Based on this SLR, a research gaps exists in proposing new approach with TM techniques to support the pre-review mapping study during the planning phase.

TABLE VII

STRATEGY AND TM TECHNIQUE(S) APPLIED TO SUPPORT SLR ACTIVITIES

Phase	SLR Activity	Sub-Activity	Level	Applied Strategy	TM Technique(s)
Planning review	1. Define Research Questions (RQs)	Identify research problems	Low	Extract key phrases and provide template for reviewers to define the RQs	IE
	2. Pre-review mapping study	Identify research gaps	High	Perform a quick informal search for relevant studies	IR, IE, IVi
	3. Develop and validate the SLR protocol	Identify fair, scientific review methodology and review objectives	Low	Offer a template and validate the user input automatically	IVi
Conducting Review	4. Search process	4.1 Search all relevant evidence, export citation list from digital database	High	Retrieve primary studies from digital libraries (e.g., meta-search, federated search). An automatic query expansion algorithm is used to optimize the users' query.	IR, clustering, classification
		4.2 Reference management	Medium	Record and maintain reference list for the identified primary studies	IR
	5. Selection Process	5.1 Citation Screening	High	Provide visualization, such as a citation map, to show the citation relationships	IE, IVi, clustering, classification
		5.2 Full text screening	High	Extract key phrases and sentences, and classify topics	IE, classification, summarization
		5.3 Decision management	Low	Record and compare the selection decisions made by different reviewers	IVi
	6. Quality Assessment	Peer evaluation for selected papers	Medium	Record and compare evaluation scores given by different reviewers	IVi
	7. Data extraction and synthesis	Present statistical results to the researcher	High	Extract the required information and present the statistical results to reviewers	IR, IE, IVi
Reporting Review	8. Write up review reports	Produce a final review report	Low	Provide a review report template, view and generate report automatically	IVi
Post-Review	9. Updating SRs	Update the SRs when new evidence becomes available	High	Notification of new publications and selection of new evidence	classification, clustering, IVi

ACKNOWLEDGMENT

This work is supported by a University of Malaya Research Grant (UMRG), Project Code: RP028C-14HTM.

REFERENCES

[1] B. Kitchenham, "Procedures for performing systematic reviews," Keele University and ESE, Nicta, UK, Australia, Tech. Rep. TR/SE-0401, 0400011T.1, July 2004.

[2] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele University and and University of Durham, UK, Tech. Rep. Ver. 2.3, EBSE-2007-01, July 2007.

[3] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *Journal of Systems and Software*, vol. 80, no. 4, pp. 571–583, 2007.

[4] M. A. Babar and H. Zhang, "Systematic literature reviews in software engineering: Preliminary results from interviews with researchers," in *Proceedings of the 3rd International Symposium on Empirical Software Engineering and Measurement*. IEEE Computer Society, 2009, pp. 346–355.

[5] M. Riaz, M. Sulayman, N. Salleh, and E. Mendes, "Experiences conducting systematic reviews from novices perspective," in *Proc. of EASE*, vol. 10, 2010, pp. 1–10.

[6] J. Carver, E. Hassler, E. Hernandes, and N. Kraft, "Identifying barriers to the systematic literature review process," in *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, Oct 2013, pp. 203–212.

[7] C. Marshall and P. Brereton, "Tools to support systematic literature reviews in software engineering: A mapping study," in *ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*. IEEE, 2013, pp. 296–299.

[8] C. Marshall, P. Brereton, and B. Kitchenham, "Tools to support systematic reviews in software engineering: A feature analysis," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. ACM, 2014, p. 13.

[9] V. Malheiros, E. Hohn, R. Pinho, and M. Mendonca, "A visual text mining approach for systematic reviews," in *International Symposium on Empirical Software Engineering and Measurement (ESEM)*. IEEE, 2007, pp. 245–254.

[10] K. R. Felizardo, E. Y. Nakagawa, D. Feitosa, R. Minghim, and J. C. Maldonado, "An approach based on visual text mining to support categorization and classification in the systematic mapping," in *Proc. of EASE*, vol. 10, 2010, pp. 1–10.

[11] K. R. Felizardo, E. Y. Nakagawa, S. G. MacDonell, and J. C. Maldonado, "A visual analysis approach to update systematic reviews," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. ACM, 2014, p. 4.

[12] H. Ramampiaro, D. Cruzes, R. Conradi, and M. Mendona, "Supporting evidence-based software engineering with collaborative information retrieval," in *6th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*. IEEE, 2010, pp. 1–5.

[13] R. Abilio, F. Morais, G. Vale, C. Oliveira, D. Pereira, and H. Costa, "Applying information retrieval techniques to detect duplicates and to rank references in the preliminary phases of systematic: Literature reviews," *CLEI Electronic Journal*, vol. 18, no. 2, pp. 3–3, 2015.

[14] A.-H. Tan *et al.*, "Text mining: The state of the art and the challenges," in *Proceedings of the PAKDD Workshop on Knowledge Discovery from Advanced Databases*, vol. 8, 1999, p. 65.

[15] R. Feldman and J. Sanger, *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press, 2007.

[16] M. W. Berry and M. Castellanos, "Survey of text mining," *Computing Reviews*, vol. 45, no. 9, p. 548, 2004.

[17] A. Hotho, A. Nürnberger, and G. Paaß, "A brief survey of text mining," *LDV Forum - GLDV Journal for Computational Linguistics and Language Technology*, vol. 20, no. 1, pp. 19–62, 2005.

[18] V. Gupta and G. S. Lehal, "A survey of text mining techniques and applications," *Journal of Emerging Technologies in Web Intelligence*, vol. 1, no. 1, pp. 60–76, 2009.

[19] C. Wohlin, "Guidelines for snowballing in systematic literature studies and a replication in software engineering," in *Proceedings of EASE*. ACM, 2014, p. 38.

[20] T. Dybå and T. Dingsøyr, "Empirical studies of agile software development: A systematic review," *Information and Software Technology*, vol. 50, no. 9, pp. 833–859, 2008.

[21] A. Nguyen-Duc, D. S. Cruzes, and R. Conradi, "The impact of global dispersion on coordination, team performance and software quality– a systematic literature review," *Information and Software Technology*, vol. 57, pp. 277–294, 2015.

[22] H. Ramampiaro, D. Cruzes, R. Conradi, and M. Mendona, "Supporting evidence-based software engineering with collaborative information re-

trieval," in *6th International Conference on Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom)*. IEEE, 2010, pp. 1–5.

[23] A. M. Fernández-Sáez, M. G. Bocco, and F. P. Romero, "Slr-tool: A tool for performing systematic literature reviews," in *5th International Conference on Software and Data Technologies (ICSOFT)*, 2010, pp. 157–166.

[24] S. Fabbri, E. Hernandes, A. Di Thommazo, A. Belgamo, A. Zamboni, and C. Silva, "Using information visualization and text mining to facilitate the conduction of systematic literature reviews," in *Enterprise Information Systems*. Springer, 2013, pp. 243–256.

[25] D. Bowes, T. Hall, and S. Beecham, "Slurp: a tool to help large complex systematic literature reviews deliver valid and rigorous results," in *Proceedings of the 2nd International Workshop on Evidential Assessment of Software Technologies*. ACM, 2012, pp. 33–36.

[26] M. Ghafari, M. Saleh, and T. Ebrahimi, "A federated search approach to facilitate systematic literature review in software engineering," *International Journal of Software Engineering and Applications*, vol. 3, no. 2, pp. 13–24, 2012.

[27] J. Torres, D. Cruzes, and L. Salvador, "Automatically locating results to support systematic reviews in software engineering," in *Workshop Latinoamericano Ingenieria de Software Experimental (ESELAW2013)*, 2014, pp. 6–19.

[28] Y. Aphinyanaphongs, I. Tsamardinos, A. Statnikov, D. Hardin, and C. F. Aliferis, "Text categorization models for high-quality article retrieval in internal medicine," *Journal of the American Medical Informatics Association*, vol. 12, no. 2, pp. 207–216, 2005.

[29] A. M. Cohen, W. R. Hersh, K. Peterson, and P.-Y. Yen, "Reducing workload in systematic review preparation using automated citation classification," *Journal of the American Medical Informatics Association*, vol. 13, no. 2, pp. 206–219, 2006.

[30] A. Kouznetsov, S. Matwin, D. Inkpen, A. H. Razavi, O. Frunza, M. Sehatkar, L. Seaward, and P. OBlenis, "Classifying biomedical abstracts using committees of classifiers and collective ranking techniques," in *Advances in Artificial Intelligence*. Springer, 2009, pp. 224–228.

[31] D. Martinez, S. Karimi,. L. Cavedon, and T. Baldwin, "Facilitating biomedical systematic reviews using ranked text retrieval and classification," in *Australasian Document Computing Symposium (ADCS)*, 2008, pp. 53–60.

[32] S. Ananiadou, B. Rea, N. Okazaki, R. Procter, and J. Thomas, "Supporting systematic reviews using text mining," *Social Science Computer Review*, vol. 27, no. 4, pp. 509–523, 2009.

[33] J. J. Yang, A. M. Cohen, and M. S. McDonagh, "Syriac: The systematic review information automated collection system a data warehouse for facilitating automated biomedical text classification," in *AMIA Annual Symposium Proceedings*, vol. 2008. American Medical Informatics Association, 2008, p. 825829.

[34] A. M. Cohen, K. Ambert, and M. McDonagh, "Cross-topic learning for work prioritization in systematic review creation and update," *Journal of the American Medical Informatics Association*, vol. 16, no. 5, pp. 690–704, 2009.

[35] S. Matwin, A. Kouznetsov, D. Inkpen, O. Frunza, and P. O'Blenis, "A new algorithm for reducing the workload of experts in performing systematic reviews," *Journal of the American Medical Informatics Association*, vol. 17, no. 4, pp. 446–453, 2010.

[36] O. Frunza, D. Inkpen, and S. Matwin, "Building systematic reviews using automatic text classification techniques," in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, 2010, pp. 303–311.

[37] B. C. Wallace, T. A. Trikalinos, J. Lau, C. Brodley, and C. H. Schmid, "Semi-automated screening of biomedical citations for systematic reviews," *BMC Bioinformatics*, vol. 11, no. 55, pp. 1–11, 2010.

[38] F. Tomassetti, G. Rizzo, A. Vetro, L. Ardito, M. Torchiano, and M. Morisio, "Linked data approach for selection process automation in systematic reviews," in *15th Annual Conference on Evaluation & Assessment in Software Engineering (EASE)*. IET, 2011, pp. 31–35.

[39] A. M. Cohen, K. Ambert, and M. McDonagh, "Studying the potential impact of automated document classification on scheduling a systematic review update," *BMC Medical Informatics and Decision Making*, vol. 12, no. 33, pp. 1–11, 2012.

[40] S. Shash and D. Mollá, "Clustering of medical publications for evidence based medicine summarisation," in *Artificial Intelligence in Medicine*, ser. Lecture Notes in Computer Science, N. Peek, R. Marín Morales, and M. Peleg, Eds. Springer Berlin Heidelberg, 2013, vol. 7885, pp. 305–309.

[41] S. Kim and J. Choi, "An svm-based high-quality article classifier for systematic reviews," *Journal of Biomedical Informatics*, vol. 47, pp. 153–159, 2014.

[42] T. Bekhuis, E. Tseytlin, K. J. Mitchell, and D. Demner-Fushman, "Feature engineering and a proposed decision-support system for systematic reviewers of medical evidence," *PloS ONE*, vol. 9, no. 1, p. e86277, 2014.

[43] J. G. Adeva, J. P. Atxa, M. U. Carrillo, and E. A. Zengotitabengoa, "Automatic text classification to support systematic reviews in medicine," *Expert Systems with Applications*, vol. 41, no. 4, pp. 1498–1508, 2014.

[44] M. Miwa, J. Thomas, A. OMara-Eves, and S. Ananiadou, "Reducing systematic review workload through certainty-based screening," *Journal of Biomedical Informatics*, vol. 51, pp. 242–253, 2014.

[45] N. R. Smalheiser, C. Lin, L. Jia, Y. Jiang, A. M. Cohen, C. Yu, J. M. Davis, C. E. Adams, M. S. McDonagh, and W. Meng, "Design and implementation of metta, a metasearch engine for biomedical literature retrieval intended for systematic reviewers," *Health Information Science and System*, vol. 2, no. 1, pp. 1–9, 2014.

[46] G. D. Mergel, M. S. Silveira, and T. S. da Silva, "A method to support search string building in systematic literature reviews through visual text mining," in *Proceedings of the 30th Annual ACM Symposium on Applied Computing*. ACM, 2015, pp. 1594–1601.

[47] R. Abilio, F. Morais, G. Vale, C. Oliveira, D. Pereira, and H. Costa, "Applying information retrieval techniques to detect duplicates and to rank references in the preliminary phases of systematic: Literature reviews," *CLEI Electronic Journal*, vol. 18, no. 2, pp. 1–24, 2015.

[48] F. Piroi, A. Lipani, M. Lupu, and A. Hanbury, "Dasyr (ir)-document analysis system for systematic reviews (in information retrieval)," in *13th International Conference on Document Analysis and Recognition (ICDAR)*. IEEE, 2015, pp. 591–595.

[49] F. Octaviano, C. Silva, and S. Fabbri, "Using the scas strategy to perform the initial selection of studies in systematic reviews: an experimental study," in *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering*. ACM, 2016, p. 25.

[50] S. Fabbri, C. Silva, E. Hernandes, F. Octaviano, A. Di Thommazo, and A. Belgamo, "Improvements in the start tool to better support the systematic review process," in *Proceedings of the 20th International Conference on Evaluation and Assessment in Software Engineering*. ACM, 2016, p. 21.

[51] G. Salton and J. Allan, "Text retrieval using the vector processing model," Nevada Univ., Las Vegas, NV (United States), Tech. Rep., 1994.

[52] A. Lopes, R. Pinho, F. V. Paulovich, and R. Minghim, "Visual text mining using association rules," *Computers & Graphics*, vol. 31, no. 3, pp. 316–326, 2007.

[53] S. N. Kim, D. Martinez, and L. Cavedon, "Automatic classification of sentences for evidence based medicine," in *Proceedings of the ACM 4th International Workshop on Data and Text Mining in Biomedical Informatics*. ACM, 2010, pp. 13–22.

[54] F. Q. Da Silva, A. L. Santos, S. Soares, A. C. C. França, C. V. Monteiro, and F. F. Maciel, "Six years of systematic literature reviews in software engineering: An updated tertiary study," *Information and Software Technology*, vol. 53, no. 9, pp. 899–913, 2011.

[55] E. Hassler, J. C. Carver, N. A. Kraft, and D. Hale, "Outcomes of a community workshop to identify and rank barriers to the systematic literature review process," in *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. ACM, 2014, p. 31.

[56] S. S.-L. Wong, N. L. Wilczynski, R. B. Haynes, R. Ramkissoonsingh, H. Team *et al.*, "Developing optimal search strategies for detecting sound clinical prediction studies in medline," in *AMIA Annual Symposium Proceedings*, vol. 2003. American Medical Informatics Association, 2003, p. 728.

[57] D. L. Sackett and W. M. Rosenberg, "The need for evidence-based medicine," *Journal of the Royal Society of Medicine*, vol. 88, no. 11, pp. 620–624, 1995.

[58] C. Wohlin, "An evidence profile for software engineering research and practice," in *Perspectives on the Future of Software Engineering*. Springer, 2013, pp. 145–157.